

Group Assignment 3

Group 10

Maria Naz

Baraa Magdy

Celia Medina Giménez

Dino Keylas

Oskar Johannes Piibar

December 22, 2025

Problem 1

Consider a three-state Markov chain with states $\{1, 2, 3\}$ and transition matrix

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 \end{pmatrix}.$$

Answer the following:

- (a) Draw the transition diagram.
- (b) Find the stationary distribution π .
- (c) Given that the chain is in state 1 at time 1, what is the probability that the chain is in state 2 at time 4?
- (d) Given that the chain is in state 1 at time 1, what is the expected time until the chain is in state 3 for the first time?
- (e) What is the period of each state?

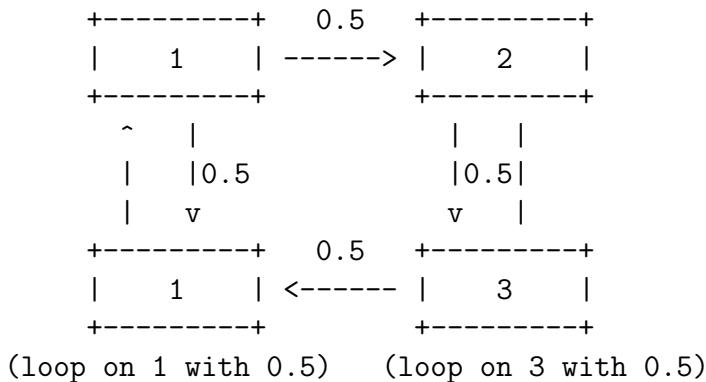
Solution

We are given a Markov chain with transition matrix

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 \end{pmatrix}.$$

(a) Transition diagram.

Below is a simple ASCII transition diagram (arrows labeled by probabilities):



(Interpretation: from $1 \rightarrow 1$ w.p. 0.5 and $1 \rightarrow 2$ w.p. 0.5; from $2 \rightarrow 1$ w.p. 0.5 and $2 \rightarrow 3$ w.p. 0.5; from $3 \rightarrow 1$ w.p. 0.5 and $3 \rightarrow 3$ w.p. 0.5.)

(b) **Stationary distribution π .**

Solve $\pi P = \pi$ with $\pi_1 + \pi_2 + \pi_3 = 1$.

From the balance equations:

$$\pi_2 = \frac{1}{2}\pi_1, \quad \pi_3 = \pi_2.$$

Normalize:

$$\pi_1 + \pi_2 + \pi_3 = \pi_1 + \frac{1}{2}\pi_1 + \frac{1}{2}\pi_1 = 2\pi_1 = 1,$$

so

$$\boxed{\pi = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)}.$$

(c) **Probability that $X_4 = 2$ given $X_1 = 1$.**

We need the $(1, 2)$ entry of P^3 . Compute (directly or by multiplication):

$$P^3 = \begin{pmatrix} 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}.$$

Hence

$$\mathbb{P}(X_4 = 2 \mid X_1 = 1) = 0.25.$$

(d) **Expected time to hit state 3 starting from 1.**

Let h_i be expected steps to first reach 3 from i ; $h_3 = 0$. Then

$$h_1 = 1 + 0.5h_1 + 0.5h_2, \quad h_2 = 1 + 0.5h_1.$$

Solving gives $h_1 = 6$ and $h_2 = 4$. Thus expected steps from 1 to first hit 3 is 6 (expected hitting index if starting at time 1 is $1 + 6 = 7$).

(e) **Period of each state.**

Since $P_{11} = 0.5 > 0$, state 1 can return in one step so has period 1. The chain is irreducible, so all states have period 1 (aperiodic).

Problem 2

Assume that we are trying to classify a binary outcome Y , i.e. our data is of the form $(X, Y) \sim F_{X,Y}$, where $Y \in \{0, 1\}$ and $X \in \mathbb{R}^d$. We have used data to train a classifier $g(X)$. We can evaluate the performance of the classifier using i.i.d. testing data $(X_1, Y_1), \dots, (X_n, Y_n)$. We are interested in estimating the following quantities:

Precision: $\mathbb{P}(Y = 1 \mid g(X) = 1)$, Recall: $\mathbb{P}(g(X) = 1 \mid Y = 1)$.

(a) Write down the empirical version of the precision and recall.

- (b) Let us now think that the variable Y denotes if a battery's health has deteriorated or not, and let X denote a bunch of constructed health indicators about the battery. If the model $g(X)$ predicts that the battery has deteriorated you need to run a test to confirm this. The cost of running the test is c when the battery is not deteriorated. On the other hand, if the battery is in fact deteriorated and the test is not run, the battery will die during use and the cost of this is d . Define a random variable representing the cost of the decision $g(X)$ and write down the formula for the expected cost in terms of the precision and recall.
- (c) Advanced question: can you produce a confidence interval for the expected cost? What about the precision and the recall?

Solution

(a) Empirical precision and recall

We know:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

Using the testing data $(X_i, Y_i)_{i=1}^n$, we define:

$$\widehat{\text{Precision}} = \frac{\#\{Y_i = 1 \wedge g(X_i) = 1\}}{\#\{g(X_i) = 1\}},$$

$$\widehat{\text{Recall}} = \frac{\#\{Y_i = 1 \wedge g(X_i) = 1\}}{\#\{Y_i = 1\}}.$$

(b) Expected cost in terms of precision and recall

If $g(X) = 1$ and $Y = 0$, we run a test unnecessarily; cost = c . If $g(X) = 0$ and $Y = 1$, we fail to detect deterioration; cost = d .

Define the random variable:

$$C = c \cdot \mathbf{1}_{\{Y=0, g(X)=1\}} + d \cdot \mathbf{1}_{\{Y=1, g(X)=0\}}.$$

To calculate the expectation:

$$\mathbb{E}[C] = \mathbb{E}(c \cdot \mathbf{1}_{\{Y=0, g(X)=1\}}) + \mathbb{E}(d \cdot \mathbf{1}_{\{Y=1, g(X)=0\}}).$$

Then (by the properties of the expectation):

$$\mathbb{E}[C] = c \mathbb{P}(Y = 0, g(X) = 1) + d \mathbb{P}(Y = 1, g(X) = 0). \quad (1)$$

But now we want to rewrite this in terms of precision and recall. Using Bayes' rule:

$$\mathbb{P}(Y = 0 \cap g(X) = 1) = \mathbb{P}(Y = 0 | g(X) = 1) \mathbb{P}(g(X) = 1)$$

$$\mathbb{P}(Y = 0 | g(X) = 1) = 1 - \mathbb{P}(Y = 1 | g(X) = 1)$$

$$\Rightarrow \mathbb{P}(Y = 0 \cap g(X) = 1) = (1 - \mathbb{P}(Y = 1 | g(X) = 1)) \mathbb{P}(g(X) = 1)$$

$$= (1 - \text{Precision}) \mathbb{P}(g(X) = 1)$$

Now, the same reasoning for recall:

$$\mathbb{P}(Y = 1, g(X) = 0) = (1 - \text{Recall}) \mathbb{P}(Y = 1).$$

Substitute into (1):

$$\boxed{\mathbb{E}[C] = c(1 - \text{Precision}) \mathbb{P}(g(X) = 1) + d(1 - \text{Recall}) \mathbb{P}(Y = 1)}$$

(c) Confidence intervals

Confidence interval for expected cost.

Each C_i takes values in $\{0, c, d\}$, hence is bounded. Let $B = \max(c, d)$. Then $C_i \in [0, B]$. Hoeffding's inequality gives, with probability at least $1 - \alpha$:

$$\left| \frac{1}{n} \sum_{i=1}^n C_i - \mathbb{E}[C] \right| \leq B \sqrt{\frac{1}{2n} \ln \left(\frac{2}{\alpha} \right)}.$$

Thus a $(1 - \alpha)$ confidence interval is:

$$\boxed{\left[\bar{C} - B \sqrt{\frac{1}{2n} \ln \left(\frac{2}{\alpha} \right)}, \quad \bar{C} + B \sqrt{\frac{1}{2n} \ln \left(\frac{2}{\alpha} \right)} \right]}$$

where \bar{C} is the sample mean.

Confidence interval for precision and recall.

Precision is the mean of $n_P = \#\{g(X_i) = 1\}$. Bernoulli trials are in $\{0, 1\}$ and have range 1.

Then we have:

$$\boxed{\delta_P = \frac{1}{\sqrt{n_P}} \sqrt{\frac{1}{2} \ln \left(\frac{2}{\alpha} \right)}}$$

$$\text{Precision} \in \left[\widehat{\text{Precision}} - \delta_P, \widehat{\text{Precision}} + \delta_P \right]$$

Similarly, recall involves $n_R = \#\{Y_i = 1\}$.

Then, we have:

$$\delta_R = \frac{1}{\sqrt{n_R}} \sqrt{\frac{1}{2} \ln \left(\frac{2}{\alpha} \right)}$$

$$\text{Recall} \in \left[\widehat{\text{Recall}} - \delta_R, \widehat{\text{Recall}} + \delta_R \right]$$

Problem 3

Let X and Y be two d -dimensional zero mean, unit variance Gaussian random vectors. Show that X and Y are nearly orthogonal by calculating their dot product. Can you for instance also bound the probability that the dot product is larger than ϵ ?

Solution

We begin by assuming that the X and Y Gaussian random vectors are independent. From the problem description and the initial assumption it can be derived that each of the components of X and Y are independent and $N(0, 1)$ distributed. This makes $X \sim N(0, I_d)$ and $Y \sim N(0, I_d)$ distributed, where I_d is the d -dimensional identity matrix.

Let the dot product of X and Y be called Z , which is defined as follows:

$$Z = \sum_{i=1}^d X_i Y_i.$$

Since we cannot be sure of the exact dot product of X and Y we will find the expected dot product, $\mathbb{E}[Z]$, which should be nearly equal to the actual one, Z .

$$\mathbb{E}[Z] = \mathbb{E} \left[\sum_{i=1}^d X_i Y_i \right]$$

Since X_i, Y_i are independent random variables, the expectation of the sum is the expectation of the elements:

$$\mathbb{E}[Z] = \mathbb{E} \left[\sum_{i=1}^d X_i Y_i \right] = \sum_{i=1}^d \mathbb{E}[X_i] \mathbb{E}[Y_i].$$

Now, each X_i and Y_i are $N(0, 1)$ distributed with $\mathbb{E}[X_i] = \mathbb{E}[Y_i] = 0$.

$$\mathbb{E}[Z] = \mathbb{E} \left[\sum_{i=1}^d X_i Y_i \right] = \sum_{i=1}^d \mathbb{E}[X_i] \mathbb{E}[Y_i] = \sum_{i=1}^d (0)(0) = 0.$$

Therefore, the expected dot product is 0 and the dot product of X and Y is nearly 0, which means that they are nearly orthogonal.

To bound the probability that the dot product is larger than ϵ , we begin by finding the variance of the dot product of X, Y :

$$V[Z] = V \left[\sum_{i=1}^d X_i Y_i \right] = \sum_{i=1}^d V[X_i Y_i], \text{ given the independence of } X_i \text{ and } Y_i.$$

Let W_i be the dot product of X_i and Y_i . Therefore,

$$V[X_i Y_i] = V[W_i] = \mathbb{E}[(W_i)^2] - (\mathbb{E}[W_i])^2 = \mathbb{E}[(W_i)^2] = \mathbb{E}[(X_i Y_i)^2]$$

, since

$$(\mathbb{E}[W_i])^2 = (\mathbb{E}[X_i Y_i])^2 = (\mathbb{E}[X_i] \mathbb{E}[Y_i])^2 = 0^2 = 0.$$

Now $V[X_i] = \mathbb{E}[(X_i)^2] - (\mathbb{E}[X_i])^2$, which can be rewritten as $\mathbb{E}[(X_i)^2] = V[X_i] + (\mathbb{E}[X_i])^2$. Each X_i and Y_i have a distribution of $N(0,1)$, therefore $\mathbb{E}[(X_i)^2] = 1 + 0^2 = 1$.

Applying this to $V[X_i Y_i]$:

$$V[X_i Y_i] = \mathbb{E}[(X_i Y_i)^2] = \mathbb{E}[(X_i)^2 (Y_i)^2] = \mathbb{E}[(X_i)^2] \mathbb{E}[(Y_i)^2] = 1 * 1 = 1.$$

We plug this in the variance equation of the dot product between X and Y :

$$V[Z] = \sum_{i=1}^d V[X_i Y_i] = \sum_{i=1}^d 1 = d.$$

Now we can use the Chebyshev's inequality to bound the probability that the dot product is larger than ϵ . The inequality for a random variable K , with mean μ and variance σ^2 is given by:

$$\mathbb{P}(|K - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Applying this to the dot product Z , with $\mu = 0$, $\sigma^2 = d$ and setting t equal to ϵ , the following equation is achieved:

$$\mathbb{P}(|Z| \geq \epsilon) \leq \frac{d}{\epsilon^2}.$$

Problem 4

Let (u_1, \dots, u_r) be $(n \times 1)$ unit-length vectors that are linearly independent, i.e., $\sum_{i=1}^r \alpha_i u_i = 0$ implies that $\alpha_i = 0$ for all i

- (a) Verify that the matrix $u_i u_i^T$ is a rank one matrix for all i . What is the null-space and range of $u_i u_i^T$?
- (b) Verify that the matrix $U = \sum_{i=1}^r u_i u_i^T$ is a rank r matrix.
- (c)
 - (i) If we perform SVD on U , are the vectors u_1, \dots, u_r the same as the right singular vectors? If not, can you give an example?
 - (ii) What if the vectors u_1, \dots, u_r are all orthogonal? In this case, what are the singular values of U ?

Solution

- (a) **Rank, Range, and Null-space of $u_i u_i^T$**

Given that $u_i \in \mathbb{R}^n$ be a unit-length vector:

$$\|u_i\| = 1 \Rightarrow u_i^T u_i = 1.$$

Consider the matrix

$$A = u_i u_i^T \in \mathbb{R}^{n \times n}.$$

Rank

One of the definition of rank is the maximum number of linearly independent vectors in Ax .

For any vector $x \in \mathbb{R}^n$:

$$Ax = u_i u_i^T x = u_i (u_i^T x).$$

- $u_i^T x$ is a scalar.
- The output is always a multiple of u_i .
- Since $u_i \neq 0$, the matrix produces exactly one independent direction.

$$\text{rank}(u_i u_i^T) = 1$$

Range (Column Space)

The range is the set of all outputs of A

From above we know that

$$(u_i u_i^T)x = u_i(u_i^T x) = u_i \times (\text{scalar})$$

So, all outputs lie on the line spanned by u_i

$$\text{Range}(A) = \{Ax : x \in \mathbb{R}^n\} = \text{span}\{u_i\}.$$

Null-space

The null-space is the set of vectors mapped to zero:

$$Ax = 0 \Rightarrow u_i u_i^T x = 0$$

In order to achieve that, and since $u_i \neq 0$, we need:

$$u_i^T x = 0$$

So:

$$\text{Null}(A) = \{x \in \mathbb{R}^n : u_i^T x = 0\}.$$

We can say that the null space of $u_i u_i^T$ is set of vectors orthogonal to u_i

- (b) **The matrix $U = \sum_{i=1}^r u_i u_i^T$ is a rank r matrix.**

1. Check the action of U on an arbitrary vector

For any $x \in \mathbb{R}^n$,

$$Ux = \sum_{i=1}^r u_i u_i^T x = \sum_{i=1}^r u_i(u_i^T x) = \sum_{i=1}^r u_i(\text{scalar}).$$

Thus, Ux is a linear combination of the vectors u_1, \dots, u_r . Therefore,

$$\text{Range}(U) \subseteq \text{span}\{u_1, \dots, u_r\}.$$

2. Characterization of the range

Since the vectors u_1, \dots, u_r are linearly independent, the space

$$\text{span}\{u_1, \dots, u_r\}$$

has dimension r . Moreover, for each $j = 1, \dots, r$,

$$Uu_j = \sum_{i=1}^r u_i(u_i^T u_j),$$

which is a nonzero vector in $\text{span}\{u_1, \dots, u_r\}$. Hence,

$$\text{Range}(U) = \text{span}\{u_1, \dots, u_r\}.$$

3. Rank of U

By definition, the rank of a matrix is the dimension of its range. Since

$$\dim(\text{Range}(U)) = \dim(\text{span}\{u_1, \dots, u_r\}) = r,$$

we conclude that

$$\boxed{\text{rank}(U) = r.}$$

□

(c i) **Investigation on whether the vectors u_1, \dots, u_r are the right singular vectors of $U = \sum_{i=1}^r u_i u_i^T$**

The matrix U is symmetric matrix since $U^T = U$.

For a symmetric matrix:

- the right singular vectors = left singular vectors
- singular vectors (left and right) are the orthonormal eigenvectors of U

Therefore, the vectors u_1, \dots, u_r coincide with the right singular vectors if and only if they are *orthonormal*, i.e.

$$u_i^T u_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

However, in the problem we are only told that the u_i are linearly independent, not orthogonal.

Therefore, in general, **they are not the right singular vectors**.

Proof with example

Let $n = 2$, and define

$$u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0.6 \\ 0.8 \end{bmatrix}.$$

Both vectors have unit length and are linearly independent.

Compute U :

$$U = u_1 u_1^T + u_2 u_2^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0.36 & 0.48 \\ 0.48 & 0.64 \end{bmatrix} = \begin{bmatrix} 1.36 & 0.48 \\ 0.48 & 0.64 \end{bmatrix}.$$

The right singular vectors of U is:

$$\begin{bmatrix} \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \end{bmatrix}$$

Therefore, **it is not equal to u_1 or u_2** .

Conclusion, ui is not right singular vectors of U

(c ii) **Singular value of U , given $u_i \dots u_r$ are all orthogonal**

If the $u_i \dots u_r$ are orthogonal and unit vectors (length = 1), it means that $u_i \dots u_r$ is orthonormal set.

Recall:

$$U = \sum_{i=1}^r u_i u_i^T$$

From the previous exercise (c i), we know that: the vectors u_1, \dots, u_r coincide with the right singular vectors if and only if they are *orthonormal*

In this case, the right (and left) singular vectors of U are $u_i \dots u_i$, so the singular value will be 1 as much as r .

Proof with example

Let $n = 2$ and $r = 2$, and define

$$u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Both vectors have unit length, linearly independent, and orthogonal to each other.

Compute U :

$$U = u_1 u_1^T + u_2 u_2^T = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Compute singular value of U :

Singular values σ_i of U are the square roots of the eigenvalues of $U^T U$.

We found that $U^T U$ is $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and the eigenvalues are $\sigma_1 = 1$ and $\sigma_2 = 1$

This is conformed that **the singular values of U is 1 as much as r** .

Problem 5

Let $X \sim \text{Uniform}(B_1)$ and define $Y = \|X\|_2$ (the Euclidean norm).

1. Find the distribution function of Y .
2. What is the distribution of $\ln(1/Y)$?
3. Calculate $\mathbb{E}[\ln(1/Y)]$, first by using the distribution function of Y and then by using the distribution function of $\ln(1/Y)$.

Solution

(a) Distribution function of Y

For $r \in [0, 1]$, CDF of Y :

$$F_Y(r) = \mathbb{P}(Y \leq r) = \mathbb{P}(\|X\|_2 \leq r).$$

Since X is uniformly distributed over B_1

$$\mathbb{P}(\|X\|_2 \leq r) = \frac{\text{Vol}(B_r)}{\text{Vol}(B_1)},$$

where $B_r = \{x \in \mathbb{R}^d : \|x\|_2 \leq r\}$. Using the scaling property of Euclidean balls in \mathbb{R}^d ,

$$\text{Vol}(B_r) = r^d \text{Vol}(B_1),$$

we obtain

$$F_Y(r) = r^d, \quad 0 \leq r \leq 1.$$

Thus,

$$F_Y(r) = \begin{cases} 0, & r < 0, \\ r^d, & 0 \leq r \leq 1, \\ 1, & r \geq 1. \end{cases}$$

(b) Distribution of $\ln(1/Y)$

Define the random variable

$$Z = \ln(1/Y).$$

For $t \geq 0$, the CDF of Z is

$$F_Z(t) = \mathbb{P}(Z \leq t) = \mathbb{P}(\ln(1/Y) \leq t).$$

Rewriting the inequality,

$$\ln(1/Y) \leq t \iff 1/Y \leq e^t \iff Y \geq e^{-t}.$$

$$F_Z(t) = \mathbb{P}(Y \geq e^{-t}) = 1 - \mathbb{P}(Y \leq e^{-t}).$$

Using the result from part (a):

$$\mathbb{P}(Y \leq e^{-t}) = (e^{-t})^d = e^{-dt},$$

$$F_Z(t) = 1 - e^{-dt}, \quad t \geq 0.$$

Then, Z follows an exponential distribution with rate parameter d :

$$\ln(1/Y) \sim \text{Exp}(d).$$

(c) Computation of $\mathbb{E}[\ln(1/Y)]$

Method 1

From part (b), since $\ln(1/Y) \sim \text{Exp}(d)$, its expectation is

$$\mathbb{E}[\ln(1/Y)] = \frac{1}{d}.$$

Method 2

From part (a), the density of Y on $(0, 1)$ is

$$f_Y(r) = dr^{d-1}.$$

$$\mathbb{E}[\ln(1/Y)] = \int_0^1 \ln(1/r) f_Y(r) dr = d \int_0^1 (-\ln r) r^{d-1} dr.$$

$$\int_0^1 r^{d-1} (-\ln r) dr = \frac{1}{d^2},$$

$$\mathbb{E}[\ln(1/Y)] = \frac{1}{d}.$$

Derivation of the integral $\int_0^1 r^{d-1}(-\ln r) dr$

Let $r = e^{-x}$. Then $x = -\ln r$ and

$$dr = -e^{-x} dx.$$

When $r = 1$, we have $x = 0$, and when $r \rightarrow 0^+$, we have $x \rightarrow \infty$. Using this change of variables, the integral becomes

$$\begin{aligned} \int_0^1 r^{d-1}(-\ln r) dr &= \int_0^\infty (e^{-x})^{d-1} x e^{-x} dx \\ &= \int_0^\infty x e^{-dx} dx. \end{aligned}$$

We now evaluate this integral by integration by parts. Let

$$u = x, \quad dv = e^{-dx} dx,$$

so that

$$du = dx, \quad v = -\frac{1}{d} e^{-dx}.$$

Then

$$\begin{aligned} \int_0^\infty x e^{-dx} dx &= \left[-\frac{x}{d} e^{-dx} \right]_0^\infty + \frac{1}{d} \int_0^\infty e^{-dx} dx \\ &= 0 + \frac{1}{d} \left[\frac{1}{d} \right] \\ &= \frac{1}{d^2}. \end{aligned}$$